DOCUMENT RESUME

ED 054 213 TM 000 788

| | |
|---|---|
| AUTHOR | Werts, Charles E.; And Others |
| TITLE | Identification and Estimation in Path Analysis with Unmeasured Variables. |
| INSTITUTION | Educational Testing Service, Princeton, N.J. |
| SPONS AGENCY | Department of Health, Education, and Welfare, Washington, D.C.; Office of Education (DHEW), Washington, D.C. |
| REPORT NO | RB-71-39 |
| PUB DATE | Jun 71 |
| GRANT | OEG-2-700033(509) |
| NOTE | 25p. |
| EDRS PRICE | EDRS Price MF-$0.65 HC-$3.29 |
| DESCRIPTORS | *Analysis of Covariance, *Correlation, Factor Analysis, Factor Structure, Goodness of Fit, *Mathematical Models, Statistical Analysis, Statistics |

ABSTRACT

A variety of path models involving unmeasured variables are formulated in terms of Joreskog's general model for the analysis of covariance structures. (Author)

RESEARCH BULLETIN

RB-71-39

IDENTIFICATION AND ESTIMATION IN PATH ANALYSIS

WITH UNMEASURED VARIABLES

Charles E. Werts, Karl G. Jöreskog

and Robert L. Linn





Educational Testing Service

Princeton, New Jersey

June 1971

Identification and Estimation in Path Analysis

with Unmeasured Variables


Abstract

A variety of path models involving unmeasured variables are formulated in terms of Jöreskog's (1970a) general model for the analysis of covariance structures.

## Identification and Estimation in Path Analysis with Unmeasured Variables*

A variety of authors (e.g., Blalock, 1969; Costner, 1969; Heise, 1969) have applied path analysis to problems involving multiple indicators of underlying constructs. An important and often algebraically complex feature of such analysis is the determination of identifiability of model parameters. The purpose of this discussion is to demonstrate how a visual inspection of the path diagram can be used to simplify the identification question and how these problems may be formulated in Jöreskog's (1970a) general model.

### I. A Single Factor Model

Consider the case of a single underlying factor $(F_1)$ with three observed measures $(X_1, X_2, \text{ and } X_3)$ as shown in Figure 1.a. The factor loadings $(\rho_{X_iF_1})$ in this model equal the standardized path coefficients $(b_1^*, b_2^*, \text{ and } b_3^*)$, given the assumption that the residuals $e_1, e_2$, and $e_3$ are independent of each other and of the factor. It is convenient, though not necessary, to assume that both measured and unmeasured variables are standardized. For heuristic purposes observed correlations will be designated with "r" and expected values of these correlations by "$\rho$". The expected correlations will differ from the corresponding observed correlations because of sampling and model specification errors.

*The research reported herein was performed pursuant to Grant No. OEG-2-700033(509) with the United States Department of Health, Education, and Welfare, and the Office of Education.

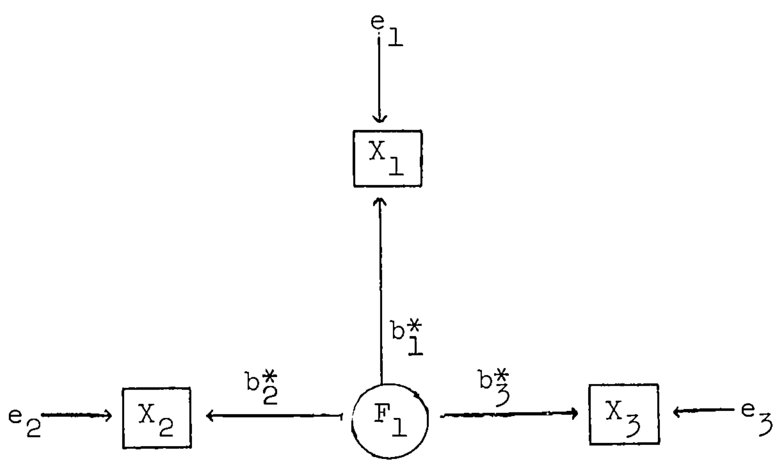



Fig. 1.a. A Single Factor Model

A path analysis of this model yields the equations:

$$\rho_{12} = b_1^* b_2^* \quad ,$$

$$\rho_{13} = b_1^* b_3^* \quad , \qquad (1)$$

and $$\rho_{23} = b_2^* b_3^* \quad .$$

Assuming nonzero correlations, equations (1) yield:

$$(b_1^*)^2 = \frac{\rho_{12}\rho_{13}}{\rho_{23}} = \rho^2_{X_1F_1} \quad ,$$

$$(b_2^*)^2 = \frac{\rho_{12}\rho_{23}}{\rho_{13}} = \rho^2_{X_2F_1} \quad \text{and} \qquad (2)$$

$$(b_3^*)^2 = \frac{\rho_{13}\rho_{23}}{\rho_{12}} = \rho^2_{X_3F_1} \quad .$$

Given only three observed measures the model is just identified, i.e., the observed and expected correlations are identical. With more than three measures

$$\rho^2_{X_iF_1} = (b_i^*)^2 = \frac{\rho_{ij}\rho_{ik}}{\rho_{jk}} \quad , \qquad \text{where } i \neq j \neq k \text{ and} \qquad (2a)$$

assuming $\rho_{jk} \neq 0$ . If there were a causal linkage (e.g., $F_1 \rightarrow I_1 \rightarrow I_2 \rightarrow X_i$) from $F_1$ to $X_i$ then $\rho_{X_iF_1}$ would be the product of the intervening path coefficients, i.e., the product of the path coefficients in the chain from $F_1$ to $X_i$ would be identified. If any loading exceeded unity, the model would be rejected. When there are $m > 3$ observed measures then the loadings will be overidentified. The number of overidentifying restrictions is simply the number of distinct correlations $m(m - 1) \div 2$ less the number $(m)$ of $\rho_{X_iF_1}$ to be estimated. Maximum likelihood or least squares estimates for over-identified models can be obtained using Jöreskog's (1970a) general method for the analysis of covariance structures. We use path analysis only to study the identifiability problem, not for estimation purposes (Hauser & Goldberger, 1970; Werts, Jöreskog, & Linn, in press).

The above analysis leads to our "rule of three": Whenever the correlations among at least three observed variables may be completely ascribed to the presence of an underlying factor, then the loadings (correlations) for each observed variable on that factor are identifiable. An important qualification is that the expected correlation between any two observed variables cannot be zero since equation (2a) would not be defined when that correlation was in the denominator. In practice, small expected correlations may lead to unstable parameter estimates, i.e., highly unreliable measures result in unreliable parameter estimates.

## II. Generalizations

The Figure 1.a. model with or without intervening, unmeasured variables going from $F_1$ to $X_i$ is too limited for most causal analyses. Our purpose in this section is to consider other causal patterns which satisfy the "rule of three," i.e., in which the observed correlations among three variables are nonzero and may be ascribed to the presence of an underlying factor. Equations (1), and therefore (2), would still hold if for one of the measures (e.g., $X_1$ ) $X_1 \to F_1$ and the residual $\theta_1$ of this regression of $F_1$ on $X_i$ were independent of the other residuals $e_2$ and $e_3$ , as shown in Figure 1.b.

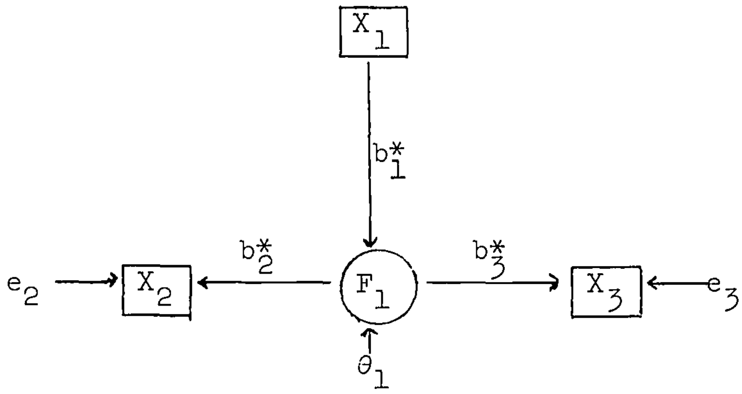



Figure 1.b.

If two observed measures influence $F_1$ , e.g., $X_1 \to F_1$ and $X_2 \to F_1$ then it is no longer true that the correlation between these measures equals the product of the corresponding path coefficients, e.g., $\rho_{12}$ would not in general equal $b^*_1 b^*_2$ .

Given that all residuals are independent, when there is an intervening variable $(I_1)$ between $X_i$ and $F_1$, the correlation between a pair of observed variables $X_i$ and $X_j$ will equal the product of the intervening path coefficients when $X_i \leftarrow I_1 \leftarrow F_1 \to X_j$ , $X_i \leftarrow I_1 \to F_1 \to X_j$ , $X_i \to I_1 \to F_1 \to X_j$ , and $X_i \leftarrow I_1 \leftarrow F_1 \leftarrow X_j$ ; but not when two arrows point towards the same variable, e.g., $X_i \to I_1 \leftarrow F_1 \to X_j$ or $X_i \to I_1 \to F_1 \leftarrow X_j$ . In general the correlation between two observed variables may be stated as the product of the intervening path coefficients whenever the causal linkage between these variables does not include a variable which is caused by two other variables, i.e., when two causal arrows point towards a variable. To identify the loadings on a factor we need to find three observed variables which are causally linked through that factor, the linkages satisfying the above criteria.

## III. Examples

A. Our first example, which corresponds to Figure 1 in Wiley and Wiley (1970), is shown in Figure 2.a.

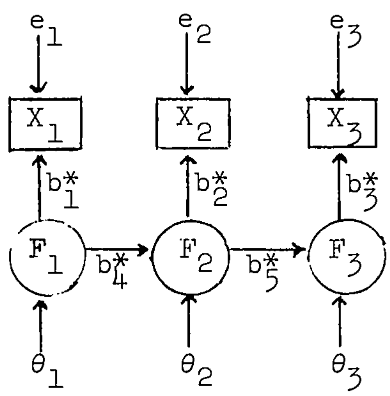



Figure 2.a.

Tracing linkages for $F_2$ :

$X_1 \leftarrow F_1 \rightarrow F_2 \rightarrow F_3 \rightarrow X_3$ ,

$X_1 \leftarrow F_1 \rightarrow F_2 \rightarrow X_2$ , and

$X_2 \leftarrow F_2 \rightarrow F_3 \rightarrow X_3$ .

Since these three linkages all include $F_2$ and satisfy the requirements of the "rule of three" we may conclude that the factor loadings $(\rho_{X_iF_2})$ , i.e., the correlations of each observed variable with $F_2$ , are identified. Thus,

$$\rho_{X_1F_2} = b_1^* b_4^* ,$$

$$\rho_{X_2F_2} = b_2^* , \text{ and} \quad (3)$$

$$\rho_{X_3F_2} = b_3^* b_5^* .$$

The factor loadings on $F_1$ are not identified because the correlation between $X_2$ and $X_3$ cannot be completely ascribed to $F_1$. Likewise the loadings on $F_3$ are not identified because the correlation between $X_1$ and $X_2$ cannot be ascribed to $F_3$. Jöreskog (1970b) shows that this model may be estimated by a single factor model with $F_2$ as the common factor and that the example may be generalized to more than three measured variables.

B. Our second example (see Figure 2.b) corresponds to Figure 4 in Costner (1969). The analysis is identical whether $F_1 \to F_2$ or $F_2 \leftarrow F_1$.

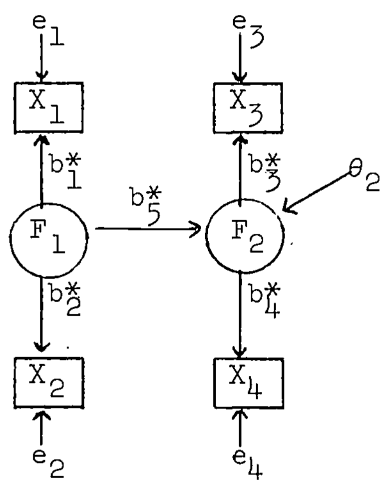



Figure 2.b.

Tracing linkages:

$$X_1 \leftarrow F_1 \rightarrow X_2 \quad , \tag{4a}$$

$$X_1 \leftarrow F_1 \rightarrow F_2 \rightarrow X_3 \quad , \tag{4b}$$

$$X_1 \leftarrow F_1 \rightarrow F_2 \rightarrow X_4 \quad , \tag{4c}$$

$$X_2 \leftarrow F_1 \rightarrow F_2 \rightarrow X_3 \quad , \tag{4d}$$

$$X_2 \leftarrow F_1 \rightarrow F_2 \rightarrow X_4 \quad , \text{ and} \tag{4e}$$

$$X_3 \leftarrow F_2 \rightarrow X_4 \quad . \tag{4f}$$

For $F_1$ the factor loadings may be identified by linkages 4a,b,d or by 4a,c,e, i.e., these loadings are overidentified and

$$\rho_{X_1F_1} = b_1^* \ ,$$

$$\rho_{X_2F_1} = b_2^* \ ,$$

$$\rho_{X_3F_1} = b_3^* b_5^* \ , \text{ and}$$

$$\rho_{X_4F_1} = b_4^* b_5^* \ .$$

The factor loadings for $F_2$ may be identified by 4b,c,f or 4d,e,f and:

$$\rho_{X_1F_2} = b_1^* b_5^* \ ,$$

$$\rho_{X_2F_2} = b_2^* b_5^* \ ,$$

$$\rho_{X_3F_2} = b_3^* \ , \text{ and}$$

$$\rho_{X_4F_2} = b_4^* \ .$$

Since $b_1^*$ and $b_2^*$ are identified, $b_5^*$ is also identified by these equations.

The analysis may be complicated by assuming $e_1$ correlated with $e_3$, in which case linkage 4b would not be valid, however the conditions of the "rule of three" would still be satisfied for $F_1$ and $F_2$ and all path coefficients and correlations between errors are (just) identified. Such a model would correspond to Figure 5.a. in Costner (1969).

C. The next example, corresponding to Figure 1 in Blalock (1963), is shown in Figure 2.c. This model is basically a variation on the model of Figure 1.b.

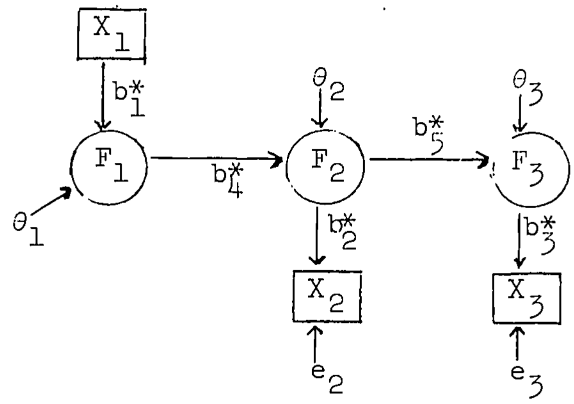



Figure 2.c.

This model differs from that in Figure 2.a. in that $X_1 \rightarrow F_1$ instead of $F_1 \rightarrow X_1$ . The linkages are:

$X_1 \rightarrow F_1 \rightarrow F_2 \rightarrow X_2$ ,

$X_1 \rightarrow F_1 \rightarrow F_2 \rightarrow F_3 \rightarrow X_3$ , and

$X_2 \leftarrow F_2 \rightarrow F_3 \rightarrow X_3$ .

Since $F_2$ is in all three linkages which satisfy the "rule of three," the factor loadings for $F_2$ are identified and

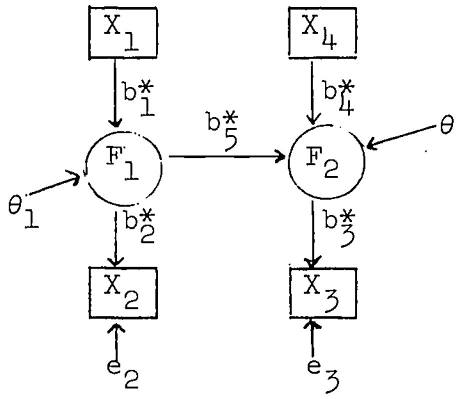
$$\rho_{X_1F_2} = b_1^*b_4^* = \sqrt{r_{12}r_{13} \div r_{23}} \quad , \tag{5a}$$

$$\rho_{X_2F_2} = b_2^* = \sqrt{r_{12}r_{23} \div r_{13}} \quad , \text{ and} \tag{5b}$$

$$\rho_{X_3F_2} = b_3^*b_5^* = \sqrt{r_{13}r_{23} \div r_{12}} \quad . \tag{5c}$$

Since $r_{12}$ cannot be ascribed to $F_3$ and $r_{23}$ cannot be ascribed to $F_1$ , the loadings on these factors are not identified. Our heuristic device would have been helpful to Blalock (1963) since he obtained the equations corresponding to the linkages shown above, but did not solve them for the equivalent of equations 5a,b, and c.

D. Our fourth example, shown in Figure 2.d., corresponds to Figure 2 in Blalock (1963).




$\theta$ = residual of $F_2$ on $X_4$ and $F_1$ .

Figure 2.d.

Tracing linkages which satisfy our rule:

$$X_1 \rightarrow F_1 \rightarrow X_2 \quad , \tag{6a}$$

$$X_1 \rightarrow F_1 \rightarrow F_2 \rightarrow X_3 \quad , \tag{6b}$$

$$X_2 \leftarrow F_1 \rightarrow F_2 \rightarrow X_3 \quad , \text{ and} \tag{6c}$$

$$X_4 \rightarrow F_2 \rightarrow X_3 \quad . \tag{6d}$$

In this model it is assumed that $X_4$ is independent of $X_1$ and $X_2$. The loadings on $F_1$ are identified by linkages 6a,b and c and therefore:

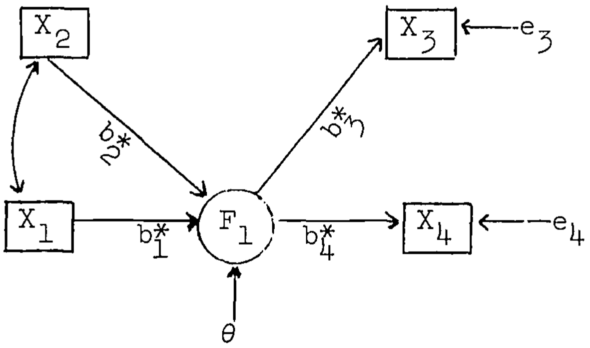
$$\rho_{X_1F_1} = b_1^* \quad , \tag{7a}$$

$$\rho_{X_2F_1} = b_2^* \quad , \text{ and} \tag{7b}$$

$$\rho_{X_3F_1} = b_3^*b_5^* \quad . \tag{7c}$$

It is not possible to find three observed variables whose linkages satisfy our rule for $F_2$ , i.e., the linkage between $X_1$ and $X_4$ has two arrows pointing at $F_2$ and the linkage between $X_1$ and $X_2$ does not include $F_2$. Since $\rho_{34} = b_3^* b_4^*$ it follows from equation (7c) that $\rho_{X_3F_1} b_4^* = \rho_{34} b_3^*$.

E. The fifth example, shown in Figure 2.e., has the special feature of two observed nonindependent variables influencing an unobserved variable. It corresponds to Figure 4 in Blalock (1969).




$\theta$ = residual of $F_1$ on $X_1$ and $X_2$ regression.

Figure 2.e.

When $X_2$ is deleted $X_1$, $X_3$, and $X_4$ form the model in Figure 1.b. from which we conclude that the correlations of $X_1$, $X_3$, and $X_4$ with $F_1$ are identified. Similarly when $X_1$ is deleted the correlations of $X_2$, $X_3$, and $X_4$ with $F_1$ are identified. Given the correlations among $X_1$, $X_2$, and $F_1$ the path coefficients $b_1^*$ and $b_2^*$ may be identified since:

$$b_1^* = \frac{\rho_{X_1F_1} - \rho_{12}\rho_{X_2F_1}}{1 - \rho_{12}^2} \quad \text{and}$$

$$b_2^* = \frac{\rho_{X_2F_1} - \rho_{12}\rho_{X_1F_1}}{1 - \rho_{12}^2} \quad .$$

F. Our last example, shown in Figure 2.f. corresponds to Figure 9.b. in Costner (1969)

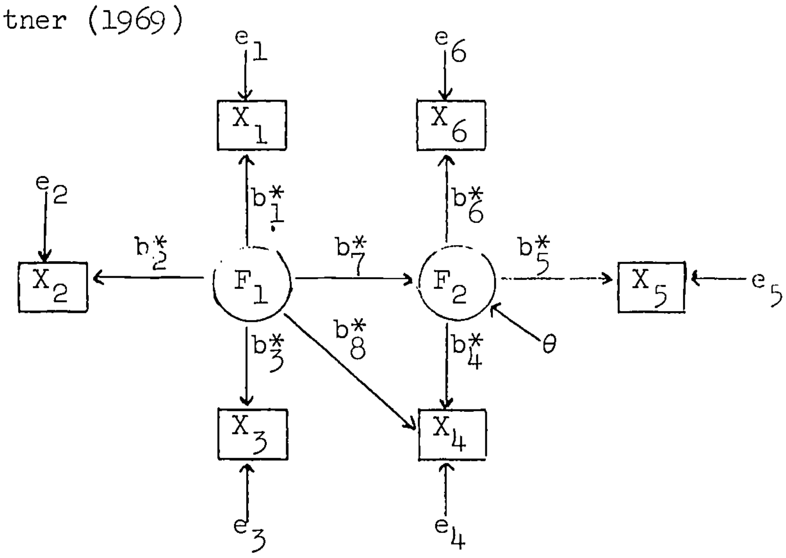



Figure 2.f.

From the analysis of the Figure 2.d. model we may deduce that when $X_4$ is excluded that $b_1^*$, $b_2^*$, $b_3^*$, $b_5^*$, $b_6^*$, and $b_7^*$ are identified. Using the variables $X_1$, $X_2$, and $X_4$ we know from our analysis of the Figure 1 model that the correlation of $X_4$ with $F_1$ ($\rho_{X_4F_1}$) is identified and similarly using $X_4$, $X_5$, and $X_6$ we know that the correlation of $X_4$ with $F_2$ ($\rho_{X_4F_2}$) is identified. Since the correlations among $F_1$, $F_2$, and $X_4$ are identified it follows that the path coefficients $b_4^*$ and $b_8^*$, which are functions of these correlations, are identified. As compared to Costner's (1969) rather complex algebraic analysis of this problem, it may be seen that we are satisfied in merely knowing that the model parameters are identified.

## IV. Estimation

Jöreskog's (1970a) general model for the analysis of covariance structures can be used to estimate the parameters for the models discussed above. Werts, Jöreskog and Linn (in press) discuss the use of Jöreskog's model from the perspective of path analysis. Use of the associated computer program (Jöreskog, Gruvaeus, & van Thillo, 1970) for the present purposes requires the investigator to specify a matrix $\Lambda$ corresponding to the factor loadings in factor analysis; a matrix $\Phi$ which is the variance-covariance matrix of the unmeasured factors, and a matrix $\Theta$ of residual variances. The matrices $B$ and $\Psi$ in Jöreskog's formula are taken as the identity and zero matrix respectively.

Consider for example the model in Figure 1 in which

$$\Lambda = \begin{bmatrix} b_1^* \\ b_2^* \\ b_3^* \end{bmatrix} \quad ,$$

$$\Phi = [1] \quad ,$$

and

$$\Theta = \begin{bmatrix} V_{e_1} & 0 & 0 \\ 0 & V_{e_2} & 0 \\ 0 & 0 & V_{e_3} \end{bmatrix} \quad .$$

Define: $\underset{\sim}{X}$ = column vector of standardized observed variables,

$\underset{\sim}{F}$ = column vector of factors, and

$\underset{\sim}{e}$ = column vector of residuals.

In matrix terminology:

$$\underset{\sim}{X} = \Lambda\underset{\sim}{F} + \underset{\sim}{e} \quad . \tag{8}$$

Equation (8) is shorthand for the path equations (all variables standardized):

$$X_1 = b_1^* F_1 + e_1 \quad ,$$

$$X_2 = b_2^* F_1 + e_2 \quad , \text{ and}$$

$$X_3 = b_3^* F_1 + e_3 \quad .$$

It can be seen that $\Lambda$ is the matrix of the coefficients of $F_1$. The parameters in the matrices specifying the model structure in Jöreskog's model are of three kinds: (1) fixed parameters that have been assigned given values; (2) constrained parameters that are unknown but equal to one or more other parameters; and (3) free parameters that are unknown and not constrained to be equal to any other parameter. In the above example the unity in $\Phi$ is a fixed parameter, whereas the $b_i^*$ in $\Lambda$ and the $V_{e_i}$ in $\Theta$ are free parameters.

The expected variance-covariance matrix $\Sigma$ for this problem is:

$$\Sigma = \Lambda\Phi\Lambda' + \Theta^2 \qquad (9)$$

where the 1 in $\Phi$ is the variance of $F_1$, for convenience standardized (i.e., equal to unity) and $\Theta^2$ is a diagonal matrix whose elements are the error variances $(V_{e_i})$. Equation (9) should be recognized as a shorthand way of expressing all the path equations relating expected model correlations to model parameters, i.e.,

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} ,$$

(where unities indicate observed variables were standardized).

Equation (9) states:

$$1 = (b_1^*)^2 + V_{e_1} \quad ,$$

$$1 = (b_2^*)^2 + V_{e_2} \quad ,$$

$$1 = (b_3^*)^2 + V_{e_3} \quad ,$$

$$\rho_{12} = b_1^* b_2^* \quad ,$$

$$\rho_{13} = b_1^* b_3^* \quad , \text{ and}$$

$$\rho_{23} = b_2^* b_3^* \quad .$$

This short description for a single model contrasts with the path analysis approach to estimation used by Costner (1969) and Blalock (1969) in the following respects:

(a) The matrix $\Sigma$ of <u>expected</u> correlations between observed variables will differ from the actually observed matrix of correlations because of sampling and/or model specification errors. Thus we do not use observed correlations in our equations as in the usual path analysis approach. Instead, Jöreskog's program attempts to minimize the difference between observed and expected variance-covariance matrices using either a least squares or maximum likelihood approach. In large samples, assuming that observed variables are distributed normally, a chi square statistic is produced which measures the overall fit of the model to the data. Another way of gauging fit is to compare the differences between the observed and expected correlations generated by the model.

(b) The degrees of freedom (df) for the $\chi^2$ measure are equal to the number of overidentifying restrictions. In path analysis this corresponds to the number of different ways the path equations may be solved for each parameter. To compute the df it is necessary to count the number of distinct elements in $\Sigma$ (i.e., $m(m + 1) \div 2$) and subtract the number of parameters to be estimated (e.g., $b_1^*, b_2^*, b_3^*, V_{e_1}, V_{e_2}$, and $V_{e_3}$) . There is no need to solve the path equations in Jöreskog's approach, although the identifiability must be known.

To analyze the model in Figure 1.b., we merely need to note that when $X_1$ and $F_1$ are standardized the regression of $X_1$ on $F_1$ equals that of $F_1$ on $X_1$ and the residuals are identical. Thus we may use the same estimation procedure for this model as for that in Figure 1.a. (where $\theta_1 = e_1$ ). Likewise the models in Figures 2.a. and 2.c. may be estimated by ignoring $F_1$ and $F_3$ and treating $X_1, X_2$, and $X_3$ as indicators of the common factor $F_2$ .

The model in Figure 2.b. with the added feature of $e_1$ and $e_3$ correlated requires special treatment. The equations are:

$$X_1 = b_1^* F_1 + e_1 \quad ,$$
$$X_2 = b_2^* F_1 + e_2 \quad ,$$
$$X_3 = b_3^* F_2 + e_3 \quad ,$$
$$X_4 = b_4^* F_2 + e_4 \quad , \text{ and}$$
$$F_2 = b_5^* F_1 + \theta_2 \quad .$$

We know that $b_5^*$ is equal to the correlation between $F_1$ and $F_2$ so there is no need to replace $F_2$ by $F_1$ and $\theta_2$ in the first four equations. To specify a correlation between $e_1$ and $e_3$ , all residuals must be treated as factors, i.e., $\underset{\sim}{F}' = (F_1, F_2, e_1, e_2, e_3, e_4)$ . The structure is:

$$\Lambda = \begin{bmatrix} b_1^* & 0 & b_{e_1}^* & 0 & 0 & 0 \\ b_2^* & 0 & 0 & b_{e_2}^* & 0 & 0 \\ 0 & b_3^* & 0 & 0 & b_{e_3}^* & 0 \\ 0 & b_4^* & 0 & 0 & 0 & b_{e_4}^* \end{bmatrix} ,$$

and

$$\Phi = \begin{bmatrix} 1 & b_5^* & 0 & 0 & 0 & 0 \\ b_5^* & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \rho_{e_1e_3} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \rho_{e_1e_3} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} .$$

In contrast to previous formulations the error variances are standardized so that the correlations between $e_1$ and $e_3$ and $F_1$ and $F_2$ are estimated directly and in $\Lambda$ the path coefficients of the observed variables on their errors $(b_{e_i}^*)$ are estimated. This model has 10 distinct elements in $\Sigma$ and 10 parameters to be estimated $(b_1^*, b_2^*, b_3^*, b_4^*, b_{e_1}^*, b_{e_2}^*, b_{e_3}^*, b_{e_4}^*, b_5^*, \rho_{e_1e_3})$, i.e., the model is just identified. The expected variance-covariance matrix $\Sigma = \Lambda\Phi\Lambda'$ , i.e., the matrix $\Theta$ is taken to be zero.

The Figure 2.d. model poses two problems: the parameters $b_3^*$ , $b_4^*$ , and $b_5^*$ are not identified and the expected correlation between $X_4$ and $X_1$ or $X_2$ is specified as zero even though the observed correlation may differ from zero presumably because of sampling fluctuations. The analysis in Section II showed that $X_4$ does not contribute to the identification of parameters, i.e., only the product $b_3^*b_5^*$ is identified with or without $X_4$. Without $X_4$ the model is that of Figure 1.b. and no purpose is served by retaining $F_2$. Assuming all variables are standardized $X_1 = b_1^*F_1 + \theta_1$ may be substituted for $F_1 = b_1^*X_1 + e_1$ as noted earlier. With $F_2$ eliminated and knowing that only the correlation of $X_3$ with $F_1$ is identified the model may be written as:

$$X_1 = b_1^*F_1 + \theta_1 \quad , \tag{10a}$$

$$X_2 = b_2^*F_1 + e_2 \quad , \text{ and} \tag{10b}$$

$$X_3 = b_3^*b_5^*F_1 + b_3^*b_4^*X_4 + e_3' \quad \text{where} \quad e_3' = b_3^*\theta + e_3 \quad . \tag{10c}$$

For convenience define $b_{35}^* = b_3^*b_5^*$ and $b_{34}^* = b_3^*b_4^*$. For computational simplicity define a new factor $x_4$ which is identical to the observed $X_4$, i.e., $X_4 = x_4$. The factors are then $\underset{\sim}{F}' = (F_1, x_4)$,

$$\Lambda = \begin{bmatrix} b_1^* & 0 \\ b_2^* & 0 \\ b_{35}^* & b_{34}^* \\ 0 & 1 \end{bmatrix} ,$$

$$\Phi = \begin{bmatrix} 1 & 0 \\ 0 & V_{x_4} \end{bmatrix} ,$$

and

$$\Theta^2 = \begin{bmatrix} V_{\theta_1} & 0 & 0 & 0 \\ 0 & V_{e_2} & 0 & 0 \\ 0 & 0 & V_{e_3'} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} .$$

In the fourth row of $\Theta^2$ the diagonal cell is zero to indicate the identity $X_4 = x_4$ without residuals. If the expected matrix $\Sigma$ is computed, i.e., $\Sigma = \Lambda\Phi\Lambda' + \Theta^2$ we find:

$$\Sigma = \begin{bmatrix} V_{X_1} & b_1^* b_2^* & b_1^* b_{35}^* & 0 \\ b_1^* b_2^* & V_{X_2} & b_2^* b_{35}^* & 0 \\ b_1^* b_{35}^* & b_2^* b_{35}^* & V_{X_3} & b_{34}^* \\ 0 & 0 & b_{34}^* & V_{X_4} \end{bmatrix} .$$

This shows that the expected correlations of $X_1$ and $X_2$ with $X_4$ are zero. This follows from the specification in $\Phi$ that $x_4$ is uncorrelated with $F_1$ .

In the analysis of the model in Figure 2.e., the correlations among $X_1$ , $X_2$ , and $F_1$ were identified first and then $b_1^*$ and $b_2^*$ identified from these correlations. The simplest estimation procedure is to estimate the correlations among $X_1$ , $X_2$ , and $F_1$ and then compute $b_1^*$ and $b_2^*$ from the estimated correlations. This problem can be handled by defining two factors $x_1 = X_1$ and $x_2 = X_2$ . The structural equations are:

$$X_1 = x_1 \quad ,$$

$$X_2 = x_2 \quad ,$$

$$X_3 = b_3^* F_1 + e_3 \quad , \text{ and}$$

$$X_4 = b_4^* F_1 + e_4 \quad .$$

The factors are $\underset{\sim}{F}' = (x_1\ ,\ x_2\ ,\ F_1)$ ,

$$\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & b_3^* \\ 0 & 0 & b_4^* \end{bmatrix} ,$$

$$\Phi = \begin{bmatrix} V_{x_1} & \rho_{x_1 x_2} & \rho_{x_1 F_1} \\ \rho_{x_1 x_2} & V_{x_2} & \rho_{x_2 F_1} \\ \rho_{x_1 F_1} & \rho_{x_2 F_1} & 1 \end{bmatrix} ,$$

and

$$\Theta^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & V_{e_3} & 0 \\ 0 & 0 & 0 & V_{e_4} \end{bmatrix} .$$

There are 10 distinct elements in $\Sigma$ and nine parameters to be estimated ($b_3^*, b_4^*, V_{x_1}, V_{x_2}, \rho_{x_1 x_2}, \rho_{X_1 F_1}, \rho_{X_2 F_1}, V_{e_3}$, and $V_{e_4}$), so that the model has one overidentifying restriction. Note that the estimated elements of $\Phi$ should be used to estimate $b_1^*$ and $b_2^*$ ($r_{X_1X_2}$ may not equal $\rho_{x_1x_2}$).

In relation to the model in Figure 2.f. Costner (1969) discussed the problem of ascertaining whether $b_8^*$ was zero and of distinguishing the $b_8^* = 0$ model from one in which errors (e.g., $e_3$ and $e_4$) were correlated. To see how this is accomplished in Jöreskog's approach, first consider the model when $b_8^* = 0$ and treating residuals as factors:

$$\underset{\sim}{X}' = (X_1, X_2, X_3, X_4, X_5, X_6) \quad ,$$

$$\underset{\sim}{F}' = (F_1, F_2, e_1, e_2, e_3, e_4, e_5, e_6) \quad ,$$

$$\Lambda = \begin{bmatrix} b_1^* & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ b_2^* & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ b_3^* & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & b_4^* & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & b_5^* & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & b_6^* & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} .$$

$$\Phi = \begin{bmatrix} 1 & b_7^* & 0 & 0 & 0 & 0 & 0 & 0 \\ b_7^* & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & V_{e_1} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & V_{e_2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & V_{e_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & V_{e_4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & V_{e_5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & V_{e_6} \end{bmatrix} ,$$

and

$$\Sigma = \Lambda\Phi\Lambda' \qquad \text{(i.e., } \Theta^2 = 0 \text{ ).}$$

Note that we have chosen not to introduce the residual $\Theta$ into the analysis because we wish to standardize both $F_1$ and $F_2$, in which case $\rho_{F_1F_2} = b_7^*$. This model is a variation of that in Figure 2.b. and all parameters are identified. There are 21 distinct elements in $\Sigma$ and 14 parameters to be estimated so that there are seven overidentifying restrictions. To test $b_8^* \neq 0$, we specify $X_4 = b_8^*F_1 + b_4^*F_2 + e_4'$, i.e., in $\Lambda$ the fourth row, first column element is left "free" instead of fixed = zero. This model has one more parameter to be estimated and therefore six overidentifying restrictions. Thus the original model is more restrictive and will therefore typically have a larger $\chi^2$. In large samples, the difference in $\chi^2$ between these two models, with degrees of freedom equal to the difference in number of restrictions, can be used to test the hypothesis that $b_8^* \neq 0$. Similarly the model with $e_3$ and $e_4$ correlated ("free") in $\Phi$ instead of independent (fixed = 0), would have six degrees of freedom and the difference in $\chi^2$ with one degree of freedom would be a test of the hypothesis that $e_3$ and $e_4$ are uncorrelated. A comparison of the $\chi^2$ for $b_8^* \neq 0$ to that for $\rho_{e_3e_4} \neq 0$ gives an indication of which is the better fitting model. Costner (1969, Figure 10) also raises the question of whether $e_1$ and $e_2$ are correlated. This hypothesis is tested by allowing the covariance between $e_1$ and $e_2$ in $\Phi$ to be "free," the change in $\chi^2$ with one degree of freedom providing the appropriate statistical test. Hypotheses involving "constrained" parameters may be tested similarly, e.g., $b_1^* = b_6^*$ (Heise, 1969) or $V_{e_1} = V_{e_6}$ (Wiley & Wiley, 1970).

It can be observed that use of Jöreskog's program requires the investigator to know the identification status of each parameter, but does not require the complex algebraic manipulations provided by Costner (1969) and Blalock (1969). It is important to recognize the essentials of each model in order to fit it into Jöreskog's general model. Jöreskog's model assumes that the observed variables are "random" rather than "fixed" but it is doubtful that most applied sociologists need to be concerned about this issue which is minor in comparison to the usual questionable validity of measures and models.

References

Blalock, H. M. Making causal inferences for unmeasured variables from correlations among indicators. American Journal of Sociology, 1963, LXIX, 53-62.

Blalock, H. M. Multiple indicators and the causal approach to measurement error. American Journal of Sociology, 1969, 75, 264-272.

Costner, H. L. Theory, deduction, and rules of correspondence. American Journal of Sociology, 1969, 75, 245-263.

Hauser, R. M., & Goldberger, A. S. The treatment of unobservable variables in path analysis. Social Systems Research Institute Series EME 7030, University of Wisconsin, Madison, Wisconsin, 1970.

Heise, D. R. Separating reliability and stability in test-retest correlation. American Sociological Review, 1969, 34, 93-101.

Jöreskog, K. G. A general method for analysis of covariance structures. Biometrika, 1970, 57, 239-251. (a)

Jöreskog, K. G. Estimation and testing of simplex models. British Journal of Mathematical and Statistical Psychology, 1970, 23, 121-145. (b)

Jöreskog, K. G., Gruvaeus, G. T., & van Thillo, M. ACOVS, a general computer program for analysis of covariance structures. Research Bulletin 70-15. Princeton, N. J.: Educational Testing Service, 1970.

Werts, C. E., Jöreskog, K. G., & Linn, R. L. Estimating the parameters of path models involving unmeasured variables. In H. M. Blalock, Jr. (Ed.), Causal models in the social sciences. Chicago: Aldine Publishing Company, in press.

Wiley, D. E., & Wiley, J. A. The estimation of measurement error in panel data. American Sociological Review, 1970, 35, 112-117.